A GENERAL FORM OF HYPOTHESIS IN THE METHOD OF FITTING CONSTANTS

by

BU-758-M January, 1982

S. R. Searle

Biometrics Unit, Cornell University, Ithaca, New York

Abstract

In the general linear model $\underset{\sim}{y} \sim N(\underset{\sim}{X}_1\underset{\sim}{\beta}_1 + \underset{\sim}{X}_2\underset{\sim}{\beta}_2 + \underset{\sim}{X}_3\underset{\sim}{\beta}_3,\ \sigma^2\underset{\sim}{I})$ the hypothesis tested when using $R(\underset{\sim}{\beta}_2|\underset{\sim}{\beta}_1)$ as the numerator sum of squares of an F-statistic is $H: \underset{\sim}{M}_1\underset{\sim}{X}_2\underset{\sim}{\beta}_2 + \underset{\sim}{M}_1\underset{\sim}{X}_2(\underset{\sim}{M}_1\underset{\sim}{X}_2)^+\underset{\sim}{X}_3\underset{\sim}{\beta}_3 = 0$ where $\underset{\sim}{M}_1 = \underset{\sim}{I} - \underset{\sim}{X}_1\underset{\sim}{X}_1^+$, and $\underset{\sim}{X}^+$ is the Moore-Penrose inverse of $\underset{\sim}{X}_1$.

1. Introduction

Consider the general linear model $\underset{\sim}{y} \sim (\underset{\sim}{X}\underset{\sim}{\beta},\ \sigma^2\underset{\sim}{I})$, namely $\underset{\sim}{y}$, a vector of order N, being normally distributed with expected value $E(\underset{\sim}{y}) = \underset{\sim}{X}\underset{\sim}{\beta}$ for $\underset{\sim}{\beta}$ of order p and $\underset{\sim}{X}$ of order $N \times p$ and rank $r_{\underset{\sim}{X}} \equiv r(\underset{\sim}{X})$, and the dispersion matrix of $\underset{\sim}{y}$ being $\sigma^2\underset{\sim}{I}$. The correct logic for testing a linear hypothesis about elements of $\underset{\sim}{\beta}$ is to formulate the hypothesis and then calculate the corresponding F-statistic using the unbiased estimator $\hat{\sigma}^2 = \underset{\sim}{y}'\underset{\sim}{M}\underset{\sim}{y}/(N - r_{\underset{\sim}{X}})$ for $\underset{\sim}{M} = \underset{\sim}{I} - \underset{\sim}{X}(\underset{\sim}{X}'\underset{\sim}{X})^-\underset{\sim}{X}' = \underset{\sim}{I} - \underset{\sim}{X}\underset{\sim}{X}^+$ where $\underset{\sim}{X}'\underset{\sim}{X}(\underset{\sim}{X}'\underset{\sim}{X})^-\underset{\sim}{X}'\underset{\sim}{X} = \underset{\sim}{X}'\underset{\sim}{X}$ and $\underset{\sim}{X}^+$ is the Moore-Penrose inverse of $\underset{\sim}{X}$. Then, on writing the hypothesis as

$$H: \underset{\sim}{K}'\underset{\sim}{\beta} = \underset{\sim}{m} \tag{1}$$

for $\underset{\sim}{K}'\underset{\sim}{\beta}$ being estimable (i.e., $\underset{\sim}{K}' = \underset{\sim}{T}'\underset{\sim}{X}$ for some $\underset{\sim}{T}'$) and $\underset{\sim}{K}'$ having full row rank, the F-statistic for testing (1) is

$$F = Q/\hat{\sigma}^2 r_{\underset{\sim}{K}} \quad \text{with} \quad Q = (\underset{\sim}{K}'\underset{\sim}{\beta}^o - \underset{\sim}{m})'[\underset{\sim}{K}'(\underset{\sim}{X}'\underset{\sim}{X})^-\underset{\sim}{K}]^{-1}(\underset{\sim}{K}'\underset{\sim}{\beta}^o - \underset{\sim}{m}) \tag{2}$$

Paper No. BU-758-M in the Biometrics Unit.

for $\beta^o = (X'X)^- X'y$. All this is well known and is detailed in many places (e.g., Searle, 1971, Chapter 5).

Unfortunately, in today's world this correct logic is often not used. Statistical computing packages spew out calculated values of sums of squares and many of them get used as numerators of F-statistics in the manner of Q in (2), but without any formal description of the hypothesis (1) accompanying them. In other words, a computed sum of squares, $y'Ay$ say, is used to calculate $F_A = y'Ay/\hat{\sigma}^2 r_A$ and the hypothesis so tested is not formally described. This note describes, for a wide class of matrices A, the hypothesis that is tested using F_A.

Knowing the hypothesis that corresponds to F_A does not justify the reverse logic of first calculating some $y'Ay$ and then ascertaining what the corresponding hypothesis is. Far from it. But the continuing existence of computed sums of squares unaccompanied by their corresponding hypotheses demands that at the very least one should know what those hypotheses are, if for no other reason than then being able to conclude from their very nature that, in numerous instances, they are hypotheses that are not worth testing. Of course, in many cases there is no difficulty. For example, consider the sum of squares for rows adjusted for columns in a row-by-column layout. With balanced data (having equal subclass numbers) the corresponding hypothesis is equality of the row effects; but with unbalanced data (having unequal subclass numbers including, perhaps, some empty cells) and a model that includes interactions, the hypothesis is not at all simple (Searle, 1971, p. 308). And if the model also includes a covariate, the exact form of the hypothesis is unknown. A general result that embraces all these cases and more is now developed.

2. The Partitioned Model

We consider the general partitioned linear model

$$E(y) = X\beta = X_1\beta_1 + X_2\beta_2 + X_3\beta_3 \ . \tag{3}$$

Attention is confined to sums of squares typified by that due to β_2 adjusted for β_1, denoted $R(\beta_2|\beta_1)$:

$$R(\beta_2|\beta_1) = R(\beta_1, \beta_2) - R(\beta_1)$$

$$= y'[X_1 \quad X_2] \begin{bmatrix} X_1'X_1 & X_1'X_2 \\ X_2'X_1 & X_2'X_2 \end{bmatrix}^- \begin{bmatrix} X_1' \\ X_2' \end{bmatrix} y - y'X_1(X_1'X_1)^- X_1'y \ . \tag{4}$$

Using the generalized inverse of a partitioned matrix available in Marsaglia and Styan (1974), straightforward algebra then reduces (4), where $M_1 = I - X_1X_1^+$, to

$$R(\beta_2|\beta_1) = y'Ay \tag{5}$$

for

$$A = M_1X_2(X_2'M_1X_2)^- X_2'M_1 = M_1X_2(M_1X_2)^+ \ . \tag{6}$$

The question then is: if $R(\beta_2|\beta_1)$ is computed, as in (4), what is the hypothesis tested by the statistic F_A for A of (6)?

3. The Hypothesis Corresponding to $R(\beta_2|\beta_1)$

It is evident from (6) that A is symmetric and idempotent. Under these conditions we have (as, for example, special cases of Theorems 1, 2 and 4 of Searle, 1971, Chapter 2)

$$E(y'Ay) = \sigma^2 r_A + \beta'X'AX\beta \ , \tag{7}$$

$$y'Ay/\sigma^2 \sim \chi^{2\prime}(r_A, \quad \beta'X'AX\beta/2\sigma^2) \ , \tag{8}$$

and

$$y'Ay \quad \text{and} \quad \hat{\sigma}^2 \text{ are independent if } (I - XX^+)A = 0 \ . \tag{9}$$

In (8), $\chi^{2\prime}(r, \lambda)$ represents the non-central χ^2 distribution with r degrees of freedom and non-centrality parameter λ. Then, from (8), we see that $y'Ay/\sigma^2$ has a central χ^2 distribution on $r_{\underset{\sim}{A}}$ degrees of freedom (denoted $\chi^2_{r_{\underset{\sim}{A}}}$) if and only if $\underset{\sim}{\beta}'\underset{\sim}{X}'\underset{\sim}{AX\beta} = 0$. On using the symmetry and idempotency of $\underset{\sim}{A}$ (and confining, very reasonably, $\underset{\sim}{AX\beta}$ to being real) this condition is equivalent to $\underset{\sim}{AX\beta} = \underset{\sim}{0}$. Hence

$$\underset{\sim}{y}'\underset{\sim}{Ay}/\sigma^2 \sim \chi^2_{r_{\underset{\sim}{A}}} \Leftrightarrow \underset{\sim}{AX\beta} = \underset{\sim}{0} \,. \tag{10}$$

Further, on writing (4) as

$$R(\underset{\sim}{\beta}_2|\underset{\sim}{\beta}_1) = \underset{\sim}{y}'\underset{\sim}{Ay} = \underset{\sim}{y}'[\underset{\sim}{X}_1\ \underset{\sim}{X}_2\ \underset{\sim}{X}_3]\left\{\begin{bmatrix} \begin{pmatrix} \underset{\sim}{X}_1'\underset{\sim}{X}_1 & \underset{\sim}{X}_1'\underset{\sim}{X}_2 \\ \underset{\sim}{X}_2'\underset{\sim}{X}_1 & \underset{\sim}{X}_2'\underset{\sim}{X}_2 \end{pmatrix}^{-} & \begin{matrix} \underset{\sim}{0} \\ \underset{\sim}{0} \end{matrix} \\ \begin{matrix} \underset{\sim}{0} & \underset{\sim}{0} \end{matrix} & \underset{\sim}{0} \end{bmatrix} - \begin{bmatrix} (\underset{\sim}{X}_1'\underset{\sim}{X}_1)^{-} & \underset{\sim}{0} & \underset{\sim}{0} \\ \underset{\sim}{0} & \underset{\sim}{0} & \underset{\sim}{0} \\ \underset{\sim}{0} & \underset{\sim}{0} & \underset{\sim}{0} \end{bmatrix}\right\}\begin{bmatrix} \underset{\sim}{X}_1' \\ \underset{\sim}{X}_2' \\ \underset{\sim}{X}_3' \end{bmatrix}\underset{\sim}{y} \tag{11}$$

it is evident that $\underset{\sim}{A} = \underset{\sim}{XCX}'$ for some $\underset{\sim}{C}$. Hence $(\underset{\sim}{I} - \underset{\sim}{XX}^+)\underset{\sim}{A} = \underset{\sim}{0}$, and so from (9), $\underset{\sim}{y}'\underset{\sim}{Ay}$ and $\hat{\sigma}^2$ are independent. Hence from (10)

$$F_A = \underset{\sim}{y}'\underset{\sim}{Ay}/\hat{\sigma}^2 r_{\underset{\sim}{A}} \quad \text{tests} \quad H: \underset{\sim}{AX\beta} = \underset{\sim}{0} \,. \tag{12}$$

Using (3) and (6), the hypothesis in (12) is

$$H: \underset{\sim}{M}_1\underset{\sim}{X}_2(\underset{\sim}{X}_2'\underset{\sim}{M}_1\underset{\sim}{X}_2)^{-}\underset{\sim}{X}_2'\underset{\sim}{M}_1(\underset{\sim}{X}_1\underset{\sim}{\beta}_1 + \underset{\sim}{X}_2\underset{\sim}{\beta}_2 + \underset{\sim}{X}_3\underset{\sim}{\beta}_3) = \underset{\sim}{0} \,. \tag{13}$$

Then, since $\underset{\sim}{M}_1\underset{\sim}{X}_1 = \underset{\sim}{0}$, and because $\underset{\sim}{M}_1$ is symmetric and idempotent and $\underset{\sim}{P}(\underset{\sim}{P}'\underset{\sim}{P})^{-}\underset{\sim}{P}'\underset{\sim}{P} = \underset{\sim}{P}$ for any $\underset{\sim}{P}$, (13) becomes

$$H: \underset{\sim}{M}_1\underset{\sim}{X}_2\underset{\sim}{\beta}_2 + \underset{\sim}{M}_1\underset{\sim}{X}_2(\underset{\sim}{M}_1\underset{\sim}{X}_2)^{+}\underset{\sim}{X}_3\underset{\sim}{\beta}_3 = \underset{\sim}{0} \,. \tag{14}$$

This, then, is the general form of the hypothesis tested by using $R(\underset{\sim}{\beta}_2|\underset{\sim}{\beta}_1)$ as the numerator of an F-statistic in the model $E(\underset{\sim}{y}) = \underset{\sim}{X}_1\underset{\sim}{\beta}_1 + \underset{\sim}{X}_2\underset{\sim}{\beta}_2 + \underset{\sim}{X}_3\underset{\sim}{\beta}_3$. It is quite general, and produces as special cases, all other combinations of model and hypothesis that are possible in this context, just five of them. All six are shown in the table, along with the values to be given to the general vector

$[\beta_1' \; \beta_2' \; \beta_3']$ to yield the special cases. For example, using $[0 \; \beta_1' \; 0]$ reduces the model to $E(y) = X_1\beta_1$, M_1 to I, M_1X_2 to X_1, $R(\beta_2|\beta_1)$ to $R(\beta_1)$ and the hypothesis to $H : X_1\beta_1 = 0$. This is line 1. Lines 2-4 and 6 are derived similarly.

4. Rank Considerations

A noticeable feature of the hypothesis $H : AX\beta = 0$ and its special cases in the table is that if written as $H : K'\beta = 0$, none of them has K' of full row rank as demanded of (1) in order for (2) to hold. Nevertheless, (2) does hold, as is now indicated. First, the symmetry of A means that $A = LL'$ for some L of full column rank r_A; and the idempotency of A means that $L'L = I$. Second, as shown following (11), $A = XCX'$ which, together with $A = LL'$ and $L'L = I$, is easily shown to imply $L = XF$ for some F. Then, because the hypothesis $H : AX\beta = 0$ of (12) is equivalent to $H : L'X\beta = 0$, we can calculate Q of (2) for this hypothesis as $Q = (L'X\beta^o)'[L'X(X'X)^-X'L]^{-1}L'X\beta^o$. The matrix inverse in Q exists because it is the inverse of $F'X'X(X'X)^-X'XF = F'X'XF = L'L = I$; and Q reduces to $y'Ay$, as it should.

A practical consequence of the hypothesis formulated as $H : AX\beta = 0$ that is important in using (14) and its special cases in the table, is that all these hypotheses are statements about N linear combinations of elements of β_2 and β_3. But only r_A of them are linearly independent. And by applying (7) to each term of (4), using $r[X(X'X)^-X'] = r(XX^+) = r_X$, in so doing, we find that $r_A = r(X_1 \,\vdots\, X_2) - r(X_1)$. This number of linearly independent linear combinations of elements of β_2 and β_3 can therefore always be used as restatement of the hypothesis (14). This is illustrated in the example that follows.

5. Example

Using $\mathbf{1}_a$ to represent an $a \times 1$ vector of ones, with $J_a = \mathbf{1}_a\mathbf{1}_a'$ and $\bar{J}_a = J_a/a$, and $D\{T\}$ for a block diagonal matrix of matrices T, we consider data from a completely randomized design of n observations in each of a classes, with a covariate represented by $\underset{\sim}{x}$. The model is

$$E(\underset{\sim}{y}) = \mu\underset{\sim}{1}_N + \left(\underset{\sim}{D}\{\underset{\sim}{1}_n\}\right)\underset{\sim}{\alpha} + b\underset{\sim}{x} ,$$

where μ represents a general mean, $\underset{\sim}{\alpha}' = [\alpha_1\ \alpha_2 \cdots \alpha_a]$ is the vector of class effects and b is the "slope" parameter for the covariate. The model (3) then has $\underset{\sim}{X}_1 = \underset{\sim}{1}_N$, $\underset{\sim}{\beta}_1 = \mu$, $\underset{\sim}{X}_2 = \underset{\sim}{D}\{\underset{\sim}{1}_n\}$, $\underset{\sim}{\beta}_2 = \underset{\sim}{\alpha}$, $\underset{\sim}{X}_3 = \underset{\sim}{x}$ and $\underset{\sim}{\beta}_3 = b$.

Suppose we seek the hypothesis corresponding to $R(\underset{\sim}{\alpha}|\mu)$. Then we have

$$\underset{\sim}{M}_1 = \underset{\sim}{I}_N - \underset{\sim}{1}_N(\underset{\sim}{1}_N'\underset{\sim}{1}_N)^{-1}\underset{\sim}{1}_N' = \underset{\sim}{I} - \underset{\sim}{\bar{J}}_N, \quad \underset{\sim}{M}_1\underset{\sim}{X}_2 = \underset{\sim}{D}\{\underset{\sim}{1}_n\} - \frac{1}{a}\underset{\sim}{J}_{N\times a},$$

$$\underset{\sim}{X}_2'\underset{\sim}{M}_1\underset{\sim}{X}_2 = n(\underset{\sim}{I}_a - \underset{\sim}{\bar{J}}_a) \quad \text{and} \quad (\underset{\sim}{X}_2'\underset{\sim}{M}_1\underset{\sim}{X}_2)^- = (1/n)(\underset{\sim}{I}_a + \lambda\underset{\sim}{J}_a) \text{ for any } \lambda .$$

It is then easily shown that

$$\underset{\sim}{M}_1\underset{\sim}{X}_2(\underset{\sim}{M}_1\underset{\sim}{X}_2)^+ = \underset{\sim}{M}_1\underset{\sim}{X}_2(\underset{\sim}{X}_2'\underset{\sim}{M}_1\underset{\sim}{X}_2)^-\underset{\sim}{X}_2'\underset{\sim}{M}_1 = \underset{\sim}{D}\{\underset{\sim}{\bar{J}}_n\} - \underset{\sim}{\bar{J}}_N$$

and so from (14) the hypothesis associated with $R(\underset{\sim}{\alpha}|\mu)$ is

$$H: \left(\underset{\sim}{D}\{\underset{\sim}{1}_n\} - \frac{1}{a}\underset{\sim}{J}_{N\times a}\right)\underset{\sim}{\alpha} + \left(\underset{\sim}{D}\{\underset{\sim}{\bar{J}}_n\} - \underset{\sim}{\bar{J}}_N\right)\underset{\sim}{x}b = \underset{\sim}{0} ,$$

i.e.,

$$H: \underset{\sim}{D}\{\underset{\sim}{1}_n\}\underset{\sim}{\alpha} - \bar{\alpha}_{.}\underset{\sim}{1}_N + b\left(\underset{\sim}{D}\{\bar{x}_{i.}\underset{\sim}{1}_n\} - \bar{x}_{..}\underset{\sim}{1}_N\right) = \underset{\sim}{0} . \tag{15}$$

This consists of N statements about the a_i's and b, but scrutiny reveals that they are n repetitions of the a statements

$$H = \alpha_i - \bar{\alpha}_{.} + b(\bar{x}_{i.} - \bar{x}_{..}) = 0 \; \forall\, i . \tag{16}$$

Since the left-hand sides of (16) sum to zero, it is equivalent to both $H: \alpha_i + b(\bar{x}_{i.} - \bar{x}_{..})$ equal $\forall$ i, and to $H: \alpha_i + b\bar{x}_{i.}$ equal $\forall$ i. Each of these is, of course, precisely as would be expected. The restatement of (15) in this form illustrates the restatement of (14) mentioned at the end of Section 4.

References

Marsaglia, George and Styan, G. P. H. (1974). Rank conditions for generalized inverses of partitioned matrices. Sankhyā 36, 437-442.

Searle, S. R. (1971). Linear Models. Wiley, New York.

Sums of Squares and Associated Hypotheses in Partitioned Linear Models.

[The general result (14) is line 5. Other lines are special cases of line 5.]

	Special value for the general vector $[\boldsymbol{\beta}_1' \ \boldsymbol{\beta}_2' \ \boldsymbol{\beta}_3']$	Model[1/] for $E(\mathbf{y})$	Sum of Squares[2/]	Associated Hypothesis[3/]
1.	$[\ \mathbf{0} \quad \boldsymbol{\beta}_1' \quad \mathbf{0}\ ]$	$\mathbf{X}_1\boldsymbol{\beta}_1$	$R(\boldsymbol{\beta}_1)$	$H: \mathbf{X}_1\boldsymbol{\beta}_1 = \mathbf{0}$
2.	$[\ \mathbf{0} \quad \boldsymbol{\beta}_1' \quad \boldsymbol{\beta}_2'\ ]$	$\mathbf{X}_1\boldsymbol{\beta}_1 + \mathbf{X}_2\boldsymbol{\beta}_2$	$R(\boldsymbol{\beta}_1)$	$H: \mathbf{X}_1\boldsymbol{\beta}_1 + \mathbf{X}_1\mathbf{X}_1^+\mathbf{X}_2\boldsymbol{\beta}_2 = \mathbf{0}$
3.	$[\ \boldsymbol{\beta}_1' \quad \boldsymbol{\beta}_2' \quad \mathbf{0}\ ]$		$R(\boldsymbol{\beta}_2 \mid \boldsymbol{\beta}_1)$	$H: \mathbf{M}_1\mathbf{X}_2\boldsymbol{\beta}_2 = \mathbf{0}$
4.	$[\ \mathbf{0} \quad \boldsymbol{\beta}_1' \quad (\boldsymbol{\beta}_2' \ \boldsymbol{\beta}_3')]$	$\mathbf{X}_1\boldsymbol{\beta}_1 + \mathbf{X}_2\boldsymbol{\beta}_2 + \mathbf{X}_3\boldsymbol{\beta}_3$	$R(\boldsymbol{\beta}_1)$	$H: \mathbf{X}_1\boldsymbol{\beta}_1 + \mathbf{X}_1\mathbf{X}_1^+(\mathbf{X}_2\boldsymbol{\beta}_2 + \mathbf{X}_3\boldsymbol{\beta}_3) = \mathbf{0}$
5.	$[\ \boldsymbol{\beta}_1' \quad \boldsymbol{\beta}_2' \quad \boldsymbol{\beta}_3'\ ]$		$R(\boldsymbol{\beta}_2 \mid \boldsymbol{\beta}_1)$	$H: \mathbf{M}_1\mathbf{X}_2\boldsymbol{\beta}_2 + \mathbf{M}_1\mathbf{X}_2(\mathbf{M}_1\mathbf{X}_2)^+\mathbf{X}_3\boldsymbol{\beta}_3 = \mathbf{0}$
6.	$[(\boldsymbol{\beta}_1' \ \boldsymbol{\beta}_2') \quad \mathbf{0} \quad \boldsymbol{\beta}_3'\ ]$		$R(\boldsymbol{\beta}_3 \mid \boldsymbol{\beta}_1, \boldsymbol{\beta}_2)$	$H: \mathbf{M}_{12}\mathbf{X}_3\boldsymbol{\beta}_3 = \mathbf{0}$

1/ In each model, $\hat{\sigma}^2 = \mathbf{y}'(\mathbf{I} - \mathbf{X}\mathbf{X}^+)\mathbf{y}/(N - r_{\mathbf{X}})$ where $\mathbf{X}$ is $\mathbf{X}_1$, $[\mathbf{X}_1 \,\vdots\, \mathbf{X}_2]$ and $[\mathbf{X}_1 \,\vdots\, \mathbf{X}_2 \,\vdots\, \mathbf{X}_3]$, respectively.

2/ The F-statistic in each case is the sum of squares divided by $s\hat{\sigma}^2$ for s being the degrees of freedom of the sum of squares, which for $R(\boldsymbol{\beta}_2 \mid \boldsymbol{\beta}_1)$ is $r_{[\mathbf{X}_1 \,\vdots\, \mathbf{X}_2]} - r_{\mathbf{X}_1}$.

3/ $\mathbf{M}_1 = \mathbf{I} - \mathbf{X}_1\mathbf{X}_1^+$ and $\mathbf{M}_{12} = \mathbf{I} - (\mathbf{X}_1 \,\vdots\, \mathbf{X}_2)(\mathbf{X}_1 \,\vdots\, \mathbf{X}_2)^+ = \mathbf{M}_1 - \mathbf{M}_1\mathbf{X}_2(\mathbf{M}_1\mathbf{X}_2)^+$.